Intrusion Detection System using Machine Learning

Synopsis Submitted To

Swami Ramanand Teerth Marathwada University, Nanded

For The Award of the Degree

Doctor of Philosophy

In the Subject of

Computer Science and Engineering

Faculty of Science and Technology

By

Gutte Prashant Haribhau

Under the Guidance of

Dr. B. R. Bombade

Professor, Department of Computer Science and Engineering

Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded (M.S.) India.

2022-2023

Contents

1 Introduction	**3**
2 Literature Survey	**4**
3 Objectives	**6**
4 Methodology	**6**
5 Proposed Plan	**7**
7 Conclusion	**7**
References	**8**

Abstract

Intrusion Detection System (IDS) is a solution to the activities of network policy violations or malicious activities in the network. A number of systems may try to prevent an intrusion attempt but this is neither required nor expected of a monitoring system. The main focus of Intrusion detection and prevention systems (IDPS) is to identify the possible incidents, logging information about them and in report attempts. In addition, organizations use IDPS for other purposes, like identifying problems with security policies, deterring individuals and documenting existing threats from infringing security policies. IDPS have become an essential addition to the securityinfrastructure of nearly every organization. Various methods can be used to detect intrusions but each one is specific to a specific method. The main goal of an intrusion detection system is to detect the attacks efficiently. Furthermore, it is equally important to detect attacks at a beginning stage in order to reduce their impacts. This research work proposed a new approach called machine learning based IDS where the anomaly dataset is measured by the ML Methods.

1 Introduction

Intrusion detection systems (IDS) are set up at a planned point within the network to examine traffic from all devices on the network. It performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known attacks. Once an attack is identified or abnormal behavior is observed, the alert can be sent to the administrator.

Intrusion Detection System (IDS) is an apache developing discipline that deals with analyzing anonymous activities occurring in network traffic. They're usually categorized as a signature-based and anomaly-based detection system. Alternative classes are network-based and host-based intrusion detection systems. Network-based IDSs have observed the malicious activities occurring in network traffic and host-based IDSs are operated on any devices and detect malicious behavior at the host. Signature-based detects malicious behavior based on previously known attacks classes. Anomaly detection creates an analysis of the normal behavior of the system and detects deviation from the model. However, having a comparatively low detection rate, anomaly detection is not used in real networks.

There is a vital necessity to deliver systems that can mechanically sight intrusion patterns and performance bottlenecks. Machine learning strategies facilitate optimizing human efforts and eventually will increase performance evaluation. The key part of such a system is to be competent in analyzing a scenario and making decisions. Machine learning and dataset processing strategies have received hefty attention for handling the drawbacks of knowledge-based detection strategies. Machine learning algorithms have worn the techniques of classification and clustering. In classification, the program should predict the foremost probable class, category or label for brand spanking new observation into one or multiple predefined categories or labels while clustering, the categories don't seem to be pre-defined throughout the learning mechanism.

Nowadays with the continuously increasing growth of network traffic, the network intrusion detection system has to analyze the network traffic within stipulated time duration and with higher detection rate. By 2020, the data generated will be 44 ZB [1]. It is mentioned that 148,809,524 packets per second need to be managed at a wire speed of 100 Gbps [1]. Hence, the network intrusion detection system has to analyze the traffic within 6.72ns [1] with an increasing level of detection rate.

In Dec 2014, Yahoo's network was hacked by a Russian hacker and therefore the info from many million accounts was purloined which incorporates names, email address, and phone numbers. Yahoo! has united to pay 35 million dollars as a penalty. With regards to all such concerns, we should have to build an intrusion detection model for analyzing such malicious activities. There are many more applications of network intrusion detection in various research areas [2-8].

The main motto of this research is to detect the divergent kind of attacks illustrated as Denial of Services (DoS), Probe, Remote to Local (R2L) and User to Root (U2R) with promising accuracy with a less computational cost. The intention of the proposed methodology is to survive the system and retain its essential services.

2 Literature Survey

There was lots of research done regarding the network intrusion detection to find anonymous activities carried by the intruders inside the network.

Ren et al. [1] bestowed a paper on network intrusion detection with reinforcement learning approach. The Large unbalanced Datasets pose challenges for ML models, as redundant and irrelevant features can hinder their effectiveness and performance of IDS can be degraded by the emergence of new network attack type in this paper MAFSIDS (Multi-Agent Feature Selection Intrusion Detection System) and DQL(Deep Q-Learning) Based IDS is used with GCN on CIC-IDS2018 nad NSL-KDD datasets for result analysis and found the selected feature subset successfully eliminates approximately 80% of redundant features compared to original feature set.

Nathan Shone [2] It uses a non-bilaterally symmetric deep autoencoder that consists of multiple auto-encoders in an exceedingly layerwise fashion which is having two asymmetrical deep belief networks having four or five shallow layers for encryption purpose and another set of four or five layers for decryption purpose with multiple hidden layers to produce the depth. The model is evaluated using the KDD Cup '99 dataset and NSL-KDD dataset with GPU enabled TensorFlow. The dataset is split into five class and 13 class classification to ensure the sustainability of the model. The model is then fit into the random forest classifier. However, the results for a few attacks categorized are anonymous.

Chuanlong Yin [10] conferred intrusion detection by using the Recurrent Neural Network (RNN). Intrusion is detected using RNNs that have a directional loop that con the previous information and applied to the present output. The experiment is performed on NSL- KDD dataset that splits into binary and five class categories. The model contains twenty hidden nodes, 0.1 as learning rate, fifty echoes and it spends 1765 seconds while not GPU acceleration. The evaluation metrics performed on WEKA tool. However, the detection rate for some attack categories is less.

Work projected in [11], conferred a replacement hybrid learning rule for accommodative network intrusion detection victimization naive Bayesian classifier and ID3 algorithm to retrieve vial features for intrusion detection. It combines associate degree, ID3 and C4.5 decision tree algorithms. The weight value is applied to every feature. The weight is decided wherever the minimum depth of the decision tree at which each feature is checked inside the tree and the weights of features that do not appear in the decision tree are allotted a value of zero. It represents that which analyzes the massive volume of network information and considers the advanced properties of attack behaviors to boost the performance of detection speed and detection accuracy. It's been successfully tested that this hybrid algorithm decreases false positives, still, as to maximize balance detection rates on the five categories of KDD Cup '99 dataset. However, the false positives of Remote to User (R2L) attack need to be improved.

Jiong Zhang [12] represents a paper on network intrusion detection systems. They have used the random forest-based classifier applied on misuse, anomaly, and hybrid-network-based intrusion detection systems. In misuse detection, the attack classes are previously known [12]. In anomaly detection, the attack is detected by using the deviation from the current activities

within the network [12]. Hybrid detection is the combination of misuse and anomaly detection. The model is evaluated using the KDD Cup '99 dataset. However, detection performance eventually decreases as the number of attack classes increases.

In [13], they have illustrated a flow-based anomaly detection in software-defined networking by building a deep neural network for an intrusion detection system and the model is built on the NSL KDD dataset. From the NSL KDD dataset, six features are extracted among the forty-one features for evaluation. A simple deep neural network with six input dimensions, two output dimensions and three hidden dimensions in which hidden layers are constructed with twelve, six and three neurons respectively is constructed; the model is classified as 2 class classification with a varying range of learning rate for obtaining better performance.

With concern to the above literature survey, we illustrate that they have achieved a higher detection rate but there are still opportunities for further enhancements. This may include the inconsistent accuracy at a higher data rate and the requirement of higher training time for a particular model. With regards to the above concerns, the model has been proposed for enhancing the detection accuracy even if the model is dealing with an increasing range of attack classes.

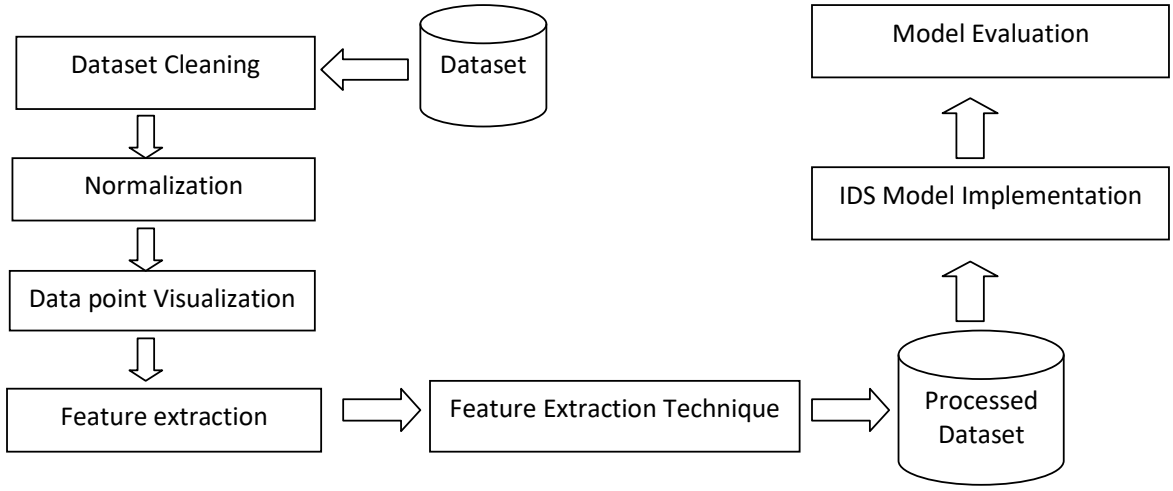
3 Objectives

The main objective of the research is to:

- Design a Machine Learning based Model to detect network intrusions.
- Recognizing features of machine learning based intrusion detection models over other models.
- Performing Accuracy Evaluation of proposed model

4 Methodology

After thorough study of currently proposed methods and their implementation, new re- search direction will be determined. This involves designing new strategies to extract prominent attacks. In our research we will be using Machine Learning based methods. Firstly,we will go through different available tools and find out the most effective results. Secondly, we will see the combination of different techniques for clear ideas if any. Finally, we will test our model and find its accuracy. We will also be focusing on finding out detailed insights and areas for further progress of research work.



a. Block diagram of methodology

5 Proposed Plan

Year	Plan
1	Literature survey and basic implementation
2	Development of Detection model and formulate results.
3	Paper writing and publishing

7 Conclusion

Intrusion Detection System (IDS) is a powerful tool that can help businesses in detecting and preventing unauthorized access to their network. By analyzing network traffic patterns, IDS can identify any suspicious activities and alert the system administrator. IDS can be a valuable addition to any organization's security infrastructure, providing insights and improving network performance.

References

[1] Kezhou Ren, Yifan Zeng, Yuanfu Zhong, *"MAFSIDS: a reinforcement learning-based intrusion detection model for multi-agent feature selection networks"*, Springer Open Journal of Big Data (2023) 10:137

[2] N. Shone, T. N. Ngoc, Vu Dinh Phai, and Qi Shi, "*A Deep Learning Approach to Network Intrusion Detection*", IEEE Transactions on Emerging Topics in Computational Intelligence, Vol. 2, NO. 1, pp. 41-50, Feb. 2018.

[3] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu, "*Deep learning for healthcare decision making with EMRs*," In the Proceedings of the 2014 IEEE International Conference on Bioinformatics.Biomed., pp. 556–559, Nov. 2014.

[4] L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, "*A deep learning-based RNNs model for an automatic security audit of short messages,*" In the Proceedings of the 16th International Symposium on Communications and Information Technologies, Qingdao, China, pp. 225–229, Sep. 2016.

[5] A. Luckow, M. Cook, N. Ashcraft, E. Weill, E. Djerekarov, and B. Vorster, "*Deep learning in theautomotive industry: Applications and tools,*" In the Proceedings of IEEE International Conference on Big Data, pp. 3759–3768. Dec. 2016.

[6] R. Polishetty, M. Roopaei, and P. Rad, "*A next-generation secure cloud-based deep learning license plate recognition for smart cities*," Proceedings of 15th IEEE International Conferenceon Machine Learning Application, Anaheim, CA, USA, pp. 286–293, Dec. 2016.

[7] H. Lee, Y. Kim, and C. O. Kim, "*A deep learning model for robust wafer fault monitoring with sensor measurement noise*," IEEE Transaction Semicond. Manuf., vol. 30, no. 1, pp. 23–31, Feb. 2017.

[8] F. Falcini, G. Lami, and A. M. Costanza, "*Deep learning in automotive software*," IEEE Softw., vol. 34, no. 3, pp. 56–63, May 2017.

[9] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "*A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology*," in Proc. IEEE EMBS Int. Conf. Biomed. Health Information, FL, USA, pp. 141–144, 2017.

[10] C. Yin, Y. Zhu, J. Fei, and X. He, *"A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks"* IEEE Access, vol. 5, pp. 21954–21961, Nov. 2017.

[11] D. Farid, N. Harbi, and Z. Rahman, "*Combining Nave Bayes and Decision Tree for Adaptive Intrusion Detection,*" International Journal of Network Security & Its Applications, Vol. 2, No.2, pp. 12-25, April 2010.

[12] J. Zhang, M. Zulkernine, A. Haque, "*Random-Forests-Based Network Intrusion Detection Systems*," IEEE Transactions on Systems, pp. 649–659, Sep 2008.

[13] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, ''*Deep learning approach for network intrusion detection in software-defined networking*,'' In the Proceedings of 2016International Conference on Wireless Networks and Mobile Communications (WINCOM), pp.258–263, Oct. 2016.

[14] B. Senthilnayaki, K. Venkatalakshmi, and A. Kannan, " *Intrusion Detection Using Optimal Genetic Feature Selection and SVM based Classifier,*" In the Proceedings of 3rd International Conference on Signal Processing Communication and Networking (ICSCN) Intrusion, pp.1–4,2015.

[15] L. Dhanabal, S. P. Shantharajah "*A Study on NSL_KDD Dataset for Intrusion Detection System Based on Classification Algorithms,*" International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, issue 6, pp. 446-452, June 2015.

[16] F. Zhang, and D. Wang, "*An Effective Feature Selection Approach for Network Intrusion Detection,*" In the Proceedings of Eighth International Conference on Networking, Architecture, and Storage, 2013.

Signature

Gutte Prashant Haribhau
Research Scholar

Signature

Dr. B. R. Bombade
Research Supervisor

ASSOCIATE PROFESSOR
Dept. of Computer Science & Engg.
S.G.G.S.Institute of Engg. & Tech.
VISHNUPURI, NANDED - 431606